

## PATENT ABSTRACTS OF JAPAN

(4)

(11)Publication number : 10-254636

(43)Date of publication of application : 25.09.1998

(51)Int.CI.

G06F 3/06

G06F 3/06

G11B 19/04

(21)Application number : 09-056234

(71)Applicant : NEC CORP

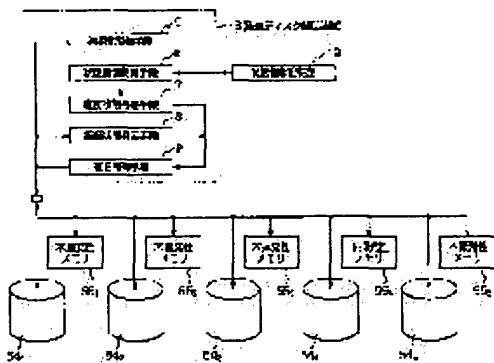
(22)Date of filing : 11.03.1997

(72)Inventor : ASANO YOSHIAKI

**(54) DISK ARRAY SYSTEM****(57)Abstract:**

**PROBLEM TO BE SOLVED:** To appropriately perform a restoration processing corresponding to the state of a magnetic disk device.

**SOLUTION:** A disk array controller 3 is provided with a state information control means 6 for rewriting state information for managing the presence/ absence of the fault generation of HDD devices 541-545 and the operation state of whether or not it is during the restoration processing corresponding to the change of the state of the HDD devices 541-55, an immediately prior state reproducing means 7 for reproducing the operation state of the respective HDD devices 541-545 immediately before the restart based on the state information at the time of restarting the HDD devices 541-545, a connection state judgement means 8 for comparing intrinsic information respectively stored in nonvolatile memories 551-555 and the HDD devices 541-545 and judging the connection state of whether or not the HDD devices 541-545 are replaced from the time immediately before the restart based on the compared result and a restoration control means 9 for controlling the continuation or start of the restoration processing based on the connection state information and operation state information.

**LEGAL STATUS**

[Date of request for examination] 11.03.1997

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 2790134

[Date of registration] 12.06.1998

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-254636

(43)公開日 平成10年(1998)9月25日

(51) Int.Cl.<sup>6</sup>  
 G 0 6 F 3/06  
 5 4 0  
 G 1 1 B 19/04

識別記号  
 3 0 4  
 5 4 0  
 5 0 1

F I  
 G 0 6 F 3/06  
 G 1 1 B 19/04

3 0 4 R  
 5 4 0  
 5 0 1 D

審査請求有 請求項の数7 O.L (全16頁)

(21)出願番号 特願平9-56234  
 (22)出願日 平成9年(1997)3月11日

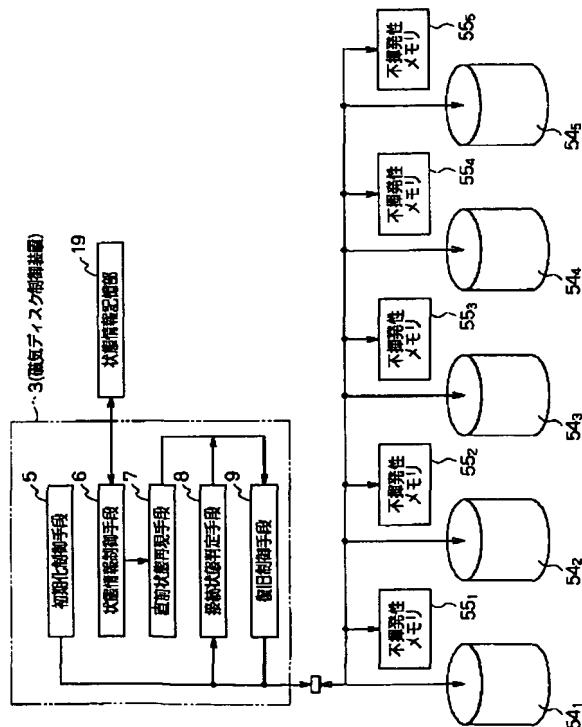
(71)出願人 000004237  
 日本電気株式会社  
 東京都港区芝五丁目7番1号  
 (72)発明者 浅野 良明  
 東京都港区芝五丁目7番1号 日本電気株式会社内  
 (74)代理人 弁理士 高橋 勇

## (54)【発明の名称】ディスクアレイシステム

## (57)【要約】

【課題】磁気ディスク装置の状態に応じて適切に復旧処理を行うこと。

【解決手段】ディスクアレイコントローラ3が、HDD装置541～545の障害発生の有無及び復旧処理中か否かの動作状態を管理する状態情報をHDD装置541～545の状態の変化に応じて書き換える状態情報制御手段6と、HDD装置541～545を再立ち上げをするときに状態情報に基づいて当該再立ち上げ直前の各HDD装置541～545の動作状態を再現する直前状態再現手段7と、不揮発性メモリ551～555及びHDD装置541～545にそれぞれ格納された固有情報を比較すると共に当該比較結果に基づいて再立ち上げ直前から当該HDD装置541～545が入れ替えられたか否かの接続状態を判定する接続状態判定手段8と、この接続状態情報と動作状態情報とに基づいて復旧処理の継続又は開始を制御する復旧制御手段9とを備えている。



## 【特許請求の範囲】

【請求項1】 上位装置から送信されたデータを冗長構成として格納する複数の磁気ディスク装置と、この複数の磁気ディスク装置を識別する固有情報を前記各磁気ディスク装置毎に記憶する不揮発性の固有情報記憶部と、この固有情報記憶部に格納された固有情報に基づいて前記磁気ディスク装置の接続状態を判定すると共に前記磁気ディスク装置に異常が発生したときに前記冗長構成に基づいて当該異常により失われたデータを復旧させる磁気ディスク制御装置とを備えたディスクアレイシステムにおいて、

前記磁気ディスク制御装置が、前記各磁気ディスク装置の初期化時に当該各磁気ディスク装置から固有情報を読み出すと共に当該各固有情報をそれぞれの前記固有情報記憶部へ格納する初期化制御手段と、前記磁気ディスク装置の障害発生の有無及び復旧処理中か否かの動作状態を管理する状態情報を前記磁気ディスク装置の状態の変化に応じて書き換える状態情報制御手段とを備え、

前記磁気ディスク制御装置に、前記状態情報制御手段によって編集された状態情報を記憶する不揮発性の状態情報記憶部を併設し、

前記磁気ディスク制御装置が、前記磁気ディスク装置を再立ち上げをするときに前記状態情報記憶部に格納された状態情報に基づいて当該再立ち上げ直前の各磁気ディスク装置の動作状態を再現する直前状態再現手段と、前記固有情報記憶部及び前記磁気ディスク装置にそれぞれ格納された固有情報を比較すると共に当該比較結果に基づいて前記再立ち上げ直前から当該磁気ディスク装置に入れ替えられたか否かの接続状態を判定する接続状態判定手段と、この接続状態判定手段によって判定された接続状態情報と前記直前状態再現手段によって再現された直前の各磁気ディスク装置の動作状態情報に基づいて前記復旧処理の継続又は開始を制御する復旧制御手段とを備えたことを特徴とするディスクアレイシステム。

【請求項2】 前記復旧制御手段が、前記直前状態再現手段によって前記直前に復旧処理中であったと判定された磁気ディスク装置が前記接続状態判定手段によって当該直前から現在まで入れ替えられていないと判定されたときには当該復旧処理を継続させる制御をする復旧継続制御機能と、前記直前状態再現手段によって前記直前に復旧処理中であったと判定された磁気ディスク装置が前記接続状態判定手段によって当該直前から現在まで入れ替えられたと判定されたときには当該復旧処理を中止させる制御をする復旧中止制御機能とを備えたことを特徴とする請求項1記載のディスクアレイシステム。

【請求項3】 前記直前状態再現手段によって前記直前に復旧処理中ではないと判定された磁気ディスク装置が前記接続状態判定手段によって当該直前から現在まで入れ替えられたと判定されたときには当該磁気ディスク装置を復旧させるか否かを前記上位装置へ問い合わせる復

旧可否問い合わせ機能を備えたことを特徴とする請求項2記載のディスクアレイシステム。

【請求項4】 前記直前状態再現手段によって前記直前に復旧中ではないと判定された磁気ディスク装置が前記接続状態判定手段によって当該直前から現在まで入れ替えられていないと判定されたときには当該磁気ディスク装置の動作テストを行うと共に当該動作テストによってエラーが生じなければ当該磁気ディスク装置を復旧させる制御をする復旧開始制御機能を備えたことを特徴とする請求項2記載のディスクアレイシステム。

【請求項5】 前記磁気ディスク制御装置が、前記固有情報を前記各固有情報記憶部に書き込んだときに当該固有情報記憶部に書き込んだ固有情報を読み出すと共に当該読み出した固有情報が書き込もうとした固有情報と一致するか否かを確認する固有情報記憶部確認手段を備え、

前記状態情報制御手段が、前記固有情報記憶部確認手段によって一致しないと判定されたときに当該固有情報の磁気ディスク装置を障害有りと判定する固有情報記憶部障害判定機能を備えたことを特徴とする請求項1記載のディスクアレイシステム。

【請求項6】 前記固有情報記憶部が、単一の不揮発性メモリを各磁気ディスク装置に割り当てた記憶領域を備え、

前記固有情報記憶部障害判定機能が、前記固有情報記憶部確認手段によって前記記憶領域に格納した固有情報が一致しないと判定されたときには前記各磁気ディスク装置の全てが障害有りと判定する機能を備えたことを特徴とする請求項5記載のディスクアレイシステム。

【請求項7】 前記磁気ディスク制御装置が、前記状態情報を前記状態情報記憶部に書き込んだときに当該状態情報記憶部に書き込んだ状態情報を読み出すと共に当該読み出した状態情報が書き込もうとした状態情報と一致するか否かを確認する状態情報記憶部確認手段を備え、前記状態情報制御手段が、前記状態情報記憶部確認手段によって一致しないと判定されたときには前記各磁気ディスク装置の全てが障害有りであると判定する状態情報記憶部障害判定機能を備えたことを特徴とする請求項1記載のディスクアレイシステム。

## 40 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、ディスクアレイシステムに係り、特に、データを冗長構成にして複数の磁気ディスク装置（HDD装置）に格納するディスクアレイシステムに関する。

## 【0002】

【従来の技術】従来、データを1台のHDD装置に格納するRAID（Redundant Arrays ofInexpensive Disk）のレベル0の磁気ディスク装置が一般的であった。

50 その後、信頼性をより向上させるため、同じデータを複

数のHDD装置に格納するRAIDのレベル1のディスクアレイシステムや、データを複数のHDD装置に分散して格納するRAIDのレベル3、4および5のディスクアレイシステムが用いられている。

【0003】RAIDのレベル0の磁気ディスク装置は、HDD装置に格納していたデータが失われた場合、そのデータを使用することはできなくなる。

【0004】RAIDのレベル1、3、4および5のディスクアレイシステムでは、データを冗長構成とするため、内蔵するHDD装置の1台に格納していたデータが失われても、ディスクアレイシステム全体としてみた場合、そのデータを復元することができる。このため、大容量の外部記憶装置が必要とされるワークステーションやネットワークサーバーなどでは、RAIDのレベル1、3、4または5を使用したディスクアレイシステムが用いられるようになってきている。

【0005】図9を用いてRAIDレベル1と呼ばれる2重化磁気ディスク装置の動作を簡単に説明する。ディスクアレイコントローラ53は、上位装置からの書き込み要求がなされると、書き込み要求がなされたデータ60を、HDD装置541、142の両方に書き込む。このような書き込み動作により、HDD装置が両方とも読み出し可能な場合には、HDD装置541とHDD装置542のデータを比較することにより、信頼性の高いデータの読み出しが可能となる。

【0006】また、一方のHDD装置のデータの読み出しが不可能になった場合でも、他方のHDD装置のデータを読み出すことにより、データを得ることができる。たとえばHDD装置541のデータが読み出し不可能となった場合には、HDD142のデータだけを読み出せば、データを得ることができる。

【0007】図10はRAIDのレベル4と呼ばれる冗長構成を用いた場合のディスクアレイシステムの構成を示す説明図である。ディスクアレイコントローラ53は、上位装置からの書き込み要求がなされると、書き込み要求がなされたデータ60を、セクタ単位に分割し、HDD装置541～144の記録領域581～584に分散して書き込む。

【0008】このとき、ディスクアレイコントローラ53は、単にデータを分散するだけでなく、対応する記録領域581～584に格納されるデータD1、D2、D3、D4の排他的論理演算を行い、その演算結果であるパリティPをHDD装置545の記録領域585に書き込む。このような書き込み動作により、データに冗長性が付与され、いずれのデータも他のデータパリティを基に再構築可能な形態で格納される。

【0009】たとえば、データD4は、データD1、D2、D3とパリティPの排他的論理演算結果と等しくなっている。このため、記録領域584が読み出し不可能となった場合には、このHDD装置の記録領域のデータ

と、パリティとを読み出してこれらの排他的論理演算を行うことにより、記録領域584を読み出すことなく、データD4を得ることができる。

【0010】また、従来のディスクアレイシステムでは、障害HDD装置を特定する情報を揮発性メモリに格納していたため、停電等によるシステムダウンが発生したとき、障害HDD装置を特定する情報が失われていた。このため、磁気ディスク装置では、障害HDD装置を特定する情報を不揮発性メモリに格納する装置が各種提案されている。

【0011】たとえば、特開平7-56694号公報には、磁気ディスク装置の状態記憶に不揮発性メモリを用いるシステムの制御方法が提案されている。

【0012】図11にこの従来のディスクアレイシステムの構成を示す。ディスクアレイシステム51は、インターフェース52と、磁気ディスク制御装置（ディスクアレイコントローラ）53と、例えば5台の磁気ディスク装置（HDD装置）541～545と、各HDD装置に対応した不揮発性メモリ551～155と、時計回路

20 16とを備えている。

【0013】インターフェース52は、上位装置61からのデータのアクセス要求（読み出しまだ書き込み要求）をディスクアレイコントローラ53に入力する。

【0014】ディスクアレイコントローラ53は、上位装置から出力された要求内容に応じてHDD装置54を制御してデータの読み出しまだ書き込みを行う手段と、不揮発性メモリ55内情報とそれぞれのディスク媒体上に設けられた管理情報記録領域17に格納された情報とに基づいて各HDD装置の状態の判定を行う手段とを備えている。時計回路16は、障害が発生した際不揮発性メモリ55の管理情報中の日付や時刻情報を書き替えるときに用いる。

【0015】この従来例の概略動作を図12及び図13を参照して説明する。この従来例では、複数の磁気ディスク装置を順次特定する変数「i」と、障害HDD装置があるときに障害HDD装置を特定するパラメータ「N\_DISK」と、障害HDD装置台数のカウント値が格納される「NEER」とを用いる。また、図11に示す各不揮発性メモリ55に格納された管理情報は配列A(i)に、

30 管理情報記憶部17に格納された管理情報は配列B

(i)に格納される。

【0016】この従来例では、1台分のデータが失われても復旧可能とする冗長構成を探るため、障害HDD装置台数のカウント値が格納される「NEER」の値が2以上の場合には、ディスクアレイシステム全体の異常とする。また、この従来例では、障害HDD装置を特定するパラメータに基づいて、障害HDD装置が新しいHDD装置と交換されたか否かを判定し、新しいHDD装置と判定されたときには、自動的に復旧処理を行う。

40 【0017】また、この従来例では、ディスクアレイコ

ントローラ53は、初期化時に、日付時刻情報を含んだ管理情報をHDD装置の管理情報記録領域57と当該HDD装置に併設された不揮発性メモリ55に格納する。また、ディスクアレイコントローラは、HDD装置に障害が発生すると、その障害HDD装置に併設された不揮発性メモリ55に障害が発生した日付時刻情報を格納する。従って、不揮発性メモリ55の内容と管理情報記録領域57の内容とを比較することで、当該HDD装置が障害ディスクであるか否かを確認することができる。

【0018】具体的には、まず、図12に示すように、配列変数A(i)、B(i)に”0”を、カウンタiに”1”、障害HDD装置台数カウンタNEERに”0”、障害HDD装置識別パラメータN<sub>DISK</sub>に”0”をセット(S201)する。

【0019】そして、i番の不揮発性メモリのテストを行い、不揮発性メモリが正常でない場合(N)には、そのHDD装置を使用不可と判断して、NEERを”1”増加させるとともに、N<sub>DISK</sub>にiをセット(S207)し、S208へ進む。

【0020】不揮発性メモリが正常である場合(S202:Y)には、その不揮発性メモリの内容を、配列変数A(i)に書き込む。次いで、i番のHDD装置の管理情報記録領域のテストを行い、管理情報記録領域が正常でない場合(S204N)には、S207へ進む。管理情報記録領域が正常である場合(S204:Y)には、その管理情報記録領域の内容をB(i)にセット(S205)する。

【0021】配列変数A(i)とB(i)が一致していないときには(S206:N) i番のHDD装置は正常なHDD装置ではないので、NEERを”1”増加させるとともに、N<sub>DISK</sub>にiをセットする(S207)。次いで、iに”1”を加算(S208)し、iがHDD装置台数より大きくなるときには(S209:Y)、S202へ戻り、次の不揮発性メモリ、管理情報記録領域のテストを行う。このような動作をiが全HDD装置台数より大きくなるまで繰り返す。

【0022】次いで、障害HDD装置台数NEERの判定を行い(S301)、NEERが”0”であるときには、全HDD装置が正常であるので、正常に立ち上げ動作を終了させる。NEERが”2”以上であるときには、エラーメッセージを出力(S302)して、再立ち上げ動作を終了する。NEERが”1”であるときには、A(N<sub>DISK</sub>)と”0”との比較を行い、A(N<sub>DISK</sub>)が”0”であるときには(S303:Y)、ステップS202で不揮発性メモリが異常と判断されその不揮発性メモリの内容が配列A(N<sub>DISK</sub>)に格納されていないため、N<sub>DISK</sub>番目の不揮発性メモリが異常であり、従って、N<sub>DISK</sub>番のHDD装置を使用不可能とする(S306)。

【0023】A(N<sub>DISK</sub>)が”0”でないときには(S303N)、N<sub>DISK</sub>番のHDD装置が使用可能な場合が

考えられるので、そのHDD装置のテストを行う(S204)。

【0024】そして、そのHDD装置を異常と判断したときには(S305:N)、そのHDD装置を使用不可とし(S306)、正常と判断したときには(S305:Y)そのHDD装置に他のHDD装置内のデータから再構築されたデータを書き込む復旧作業を行う(S307)。復旧作業が完了したならば、そのHDD装置の管理情報記録領域にA(N<sub>DISK</sub>)を書き込み、再立ち上げ動作を完了する。

#### 【0025】

【発明が解決しようとする課題】しかしながら、この従来例では、障害HDD装置を取り替えるときに、間違えて、データがすでに格納されているHDD装置を新しいHDD装置として接続した場合には、なんらエラーが検出されず、このため、データがすでに格納されているHDD装置へ復旧データが書き込まれてしまい、元のデータが失われてしまう、という不都合があった。

【0026】すなわち、図11に示した構成を持つディスクアレイシステムが複数あり、障害HDD装置が複数台になった場合、図12及び図13に示す手法では、障害HDD装置交換後の再立ち上げ時に誤動作を生ずる。

【0027】たとえば、図14に示すように図11の磁気ディスク装置が2台あり、交換するHDD装置が2台ある場合で、ディスクアレイシステム51aのHDD装置54a<sub>1</sub>とディスクアレイシステム51bのHDD装置54b<sub>2</sub>が障害HDD装置になった場合を考える。

【0028】ディスクアレイシステム51aとディスクアレイシステム51bはそれぞれ独立して動作しているため、HDD装置54a<sub>1</sub>～54a<sub>5</sub>と54b<sub>1</sub>～54b<sub>5</sub>に書き込まれているデータは異なるものである。

【0029】障害HDD装置54a<sub>1</sub>および54b<sub>2</sub>を新しいHDD装置54c<sub>1</sub>および14c<sub>2</sub>に交換する際に、ディスクアレイシステム51aの障害HDD装置54a<sub>1</sub>ではなく、正常に動作していたHDD装置54a<sub>2</sub>を取り外し、新しいHDD装置54c<sub>1</sub>と交換し、ディスクアレイシステム51bの障害HDD装置54b<sub>2</sub>を取り外し、先に取り外したディスクアレイシステム51aで正常に動作していたHDD装置54a<sub>2</sub>を取り付け、両方の磁気ディスク装置の交換作業が終了後同時に再起動したとする。

【0030】すなわち、障害のない磁気ディスク装置を新しい交換用の磁気ディスク装置に変更したり、障害のある磁気ディスク装置を交換用の磁気ディスク装置と間違って交換したとする。

【0031】この場合、ディスクアレイシステム51aは、障害HDD装置54a<sub>1</sub>と新しいHDD装置54c<sub>1</sub>(HDD装置54a<sub>2</sub>があつた場所)が障害HDD装置となるため、使用不可能な状態になる。一方、ディスクアレイシステム51bは障害HDD装置54b<sub>2</sub>が正

常動作するHDD装置54a<sub>2</sub>となるので、復旧動作を開始することになる。

【0032】ディスクアレイシステム51bの復旧動作を行うとディスクアレイシステム51aのHDD装置54a<sub>2</sub>のデータがなくなることになるので、ディスクアレイシステム51aはHDD装置54a<sub>1</sub>と54a<sub>2</sub>の2つのデータがなくなることになり、復旧作業が行えなくなる。

【0033】すなわち、一方のディスクアレイシステムから正常な磁気ディスク装置を取り外し、他方のディスクアレイシステムに接続して復旧作業を行うと、一方のディスクアレイシステムは障害復旧が行えなくなる、という不都合があった。

【0034】このような問題は、HDD装置の交換接続作業を磁気ディスク装置の使用者が間違いなく行えば、生じない問題である。しかし、磁気ディスク装置がデータを冗長構成とすることにより、記憶装置としての信頼性を向上させた装置である以上、このような人為的なミスにも対応し得る装置であることが望まれる。

【0035】このように、従来例では、復旧処理をすることでデータが失われてしまう可能性が考慮されておらず、従って、適切な局面での復旧処理を行うことができない、という不都合があった。

【0036】また、不揮発性メモリ551～154を1つの不揮発性メモリにして、管理情報を格納する領域を区別することにより、各HDD装置に対応させることができになっているが、不揮発性メモリを1つにした場合は、図12および図13に示すシステムの制御方法の不揮発性メモリテストは、それぞれの不揮発性メモリの管理情報記憶領域のテストを行うことになり、不揮発性メモリ自体の障害を検出できない場合がある。

【0037】たとえば、HDD装置に障害が発生して不揮発性メモリの管理情報記録領域を書き替えるとき、今まで正常に動作していた不揮発性メモリが書き込み障害を起こし、データが書き替えられなかつたか、または書き替えたデータが書き替える前と同じであったとする。

【0038】その後、障害に気づかずに入力をすると不揮発性メモリに書き込み障害を起こした管理情報記録領域のテストは正常に終了し、装置は障害がなかつたようになる。また、不揮発性メモリの書き込み障害を起こした管理情報記録領域のテストがエラーとなつても他の管理情報記録領域は正常に見える。

【0039】これは、1つの不揮発性メモリを使用した磁気ディスク装置で不揮発性メモリの障害が発生した際に、HDD装置の障害を検出することができない場合や、障害が発生した領域は使用できないがその他の領域は使用できる場合などが生じることになり、正しい磁気ディスク装置の状態を知ることができなくなる。

【0040】この問題は、回路上の様々な障害が原因となって生じることが考えられる。そのため、このような

問題が生じ際に、障害を検出する機構を持つ装置が望まれる。

#### 【0041】

【発明の目的】本発明は、係る従来例の有する不都合を改善し、特に、磁気ディスク装置の状態に応じて適切に復旧処理を行うことのできるディスクアレイシステムを提供することを、その目的とする。

【0042】具体的には、各HDD装置の状態を立ち上げ時に自動検出するとともに不揮発性メモリの障害も検出することにより、システムダウン後の再立ち上げ時に誤動作を生じることがくなるようにして、システムの信頼性の向上を図る。

【0043】また、HDD装置の接続ミスも検出することにより、人為的ミスによりデータが失われることがなるようにしてシステムの信頼性の向上を図る。

【0044】さらに、障害HDD装置が交換された時に、人手介入により復旧を開始することにより、交換時のミスによりデータが失われることを防げるようにして、復旧を開始した後は、システムがダウンしても再立ち上げ時に自動検出して復旧を開始することにより、上位装置を介して入力する必要がなくなるようにしてシステムの保守性の向上を図る。

#### 【0045】

【課題を解決するための手段】そこで、本発明では、上位装置から送信されたデータを冗長構成として格納する複数の磁気ディスク装置と、この複数の磁気ディスク装置を識別する固有情報を各磁気ディスク装置毎に記憶する不揮発性の固有情報記憶部と、この固有情報記憶部に格納された固有情報に基づいて磁気ディスク装置の接続状態を判定すると共に磁気ディスク装置に異常が発生したときに冗長構成に基づいて当該異常により失われたデータを復旧させる磁気ディスク制御装置とを備えている。しかも、磁気ディスク制御装置が、各磁気ディスク装置の初期化時に当該各磁気ディスク装置から固有情報を読み出すと共に当該各固有情報をそれぞれの固有情報記憶部に格納する初期化制御手段と、磁気ディスク装置の障害発生の有無及び復旧処理中か否かの動作状態を管理する状態情報を磁気ディスク装置の状態の変化に応じて書き換える状態情報制御手段とを備えている。さら

に、磁気ディスク制御装置に、状態情報制御手段によって編集された状態情報を記憶する不揮発性の状態情報記憶部を併設している。そして、磁気ディスク制御装置が、磁気ディスク装置を再立ち上げをするときに状態情報記憶部に格納された状態情報に基づいて当該再立ち上げ直前の各磁気ディスク装置の動作状態を再現する直前状態再現手段と、固有情報記憶部及び磁気ディスク装置にそれぞれ格納された固有情報を比較すると共に当該比較結果に基づいて再立ち上げ直前から当該磁気ディスク装置が入れ替えられたか否かの接続状態を判定する接続状態判定手段と、この接続状態判定手段によって判定さ

れた接続状態情報と直前状態再現手段によって再現された直前の各磁気ディスク装置の動作状態情報に基づいて復旧処理の継続又は開始を制御する復旧制御手段とを備えた、という構成を探っている。これにより前述した目的を達成しようとするものである。

【0046】本発明によるディスクアレイシステムは、データを冗長構成として各磁気ディスク装置に格納している。このため、例えば1台の磁気ディスク装置に障害が発生し、データの読み出しができなくなったとしても、他の磁気ディスク装置に格納されたデータから当該障害が発生した磁気ディスク装置に格納されていたデータを復旧することができる。

【0047】また、本発明では、初期化制御手段が、各磁気ディスク装置の初期化時に当該各磁気ディスク装置から固有情報を読み出すと共に当該各固有情報をそれぞれの固有情報記憶部へ格納する。この固有情報は、各磁気ディスク装置に必ず付されている情報を用いる。例えば、ベンダー名と各磁気ディスク装置の識別番号（シリアルナンバー）を用いるとよい。このため、磁気ディスク制御装置が独自に固有情報を付する場合と比較して、より確実な各磁気ディスク装置の識別が行われる。従って、接続状態判定手段は、固有情報の比較により、電源の投入や障害発生後の再立ち上げのときに、磁気ディスク装置が交換されたか否かを確実に判定する。

【0048】さらに、本発明では、状態情報制御手段が、磁気ディスク装置の障害発生の有無及び復旧処理中か否かの動作状態を管理する状態情報を磁気ディスク装置の状態の変化に応じて書き換え、復旧制御手段が、接続状態判定手段によって判定された接続状態情報と直前状態再現手段によって再現された直前の各磁気ディスク装置の動作状態情報に基づいて復旧処理の継続又は開始を制御する。従って、立ち上げ前の直前の状態と、現在の接続状態とに応じて、復旧処理を行うか否かを定める。このため、人為的な接続ミスがあった場合には、これを確実に検出し、復旧処理を中止する等の処理が可能となる。

#### 【0049】

【発明の実施の形態】図1は本発明の一実施形態の構成を示すブロック図である。本実施形態によるディスクアレイシステムは、図1に示すように、上位装置から送信されたデータを冗長構成として格納する複数の磁気ディスク装置（HDD装置）541～545と、この複数のHDD装置541～545を識別する固有情報を各HDD装置毎に記憶する不揮発性の不揮発性メモリ（不揮発性メモリ）551～555と、この不揮発性メモリ551～555に格納された固有情報に基づいてHDD装置541～545の接続状態を判定すると共にHDD装置541～545に異常が発生したときに冗長構成に基づいて当該異常により失われたデータを復旧させる磁気ディスク制御装置（ディスクアレイコントローラ）3とを備えてい

る。

【0050】しかも、ディスクアレイコントローラ3が、HDD装置541～545の初期化時に当該各HDD装置541～545から固有情報を読み出すと共に当該各固有情報をそれぞれの不揮発性メモリ551～555へ格納する初期化制御手段5と、HDD装置541～545の障害発生の有無及び復旧処理中か否かの動作状態を管理する状態情報をHDD装置541～545の状態の変化に応じて書き換える状態情報制御手段6とを備えている。

【0051】さらに、ディスクアレイコントローラ3に、状態情報制御手段6によって編集された状態情報を記憶する不揮発性の状態情報記憶部19を併設している。

【0052】そして、ディスクアレイコントローラ3が、HDD装置541～545を再立ち上げをするときに状態情報記憶部19に格納された状態情報に基づいて当該再立ち上げ直前の各HDD装置541～545の動作状態を再現する直前状態再現手段7と、不揮発性メモリ551～555及びHDD装置541～545にそれぞれ格納

された固有情報を比較すると共に当該比較結果に基づいて再立ち上げ直前から当該HDD装置541～545が入れ替えられたか否かの接続状態を判定する接続状態判定手段8と、この接続状態判定手段8によって判定された接続状態情報と直前状態再現手段7によって再現された直前の各HDD装置541～545の動作状態情報に基づいて復旧処理の継続又は開始を制御する復旧制御手段9とを備えている。

【0053】初期化制御手段5が、HDD装置に固有な情報を不揮発性メモリ55に格納するため、接続状態判定手段8は、不揮発性メモリ551～555及びHDD装置541～545にそれぞれ格納された固有情報を比較すると共に当該比較結果に基づいて再立ち上げ直前から当該HDD装置541～545が入れ替えられたか否かの接続状態を判定することができる。すなわち、HDD装置が交換された場合、不揮発性メモリに格納された固有情報とその交換されたHDD装置の固有情報は必ず異なり、一方、交換されていない場合には必ず一致するため、接続状態判定手段8は、現在のHDD装置の接続状態を正確に判定することができる。

【0054】さらに、状態情報制御手段6が、HDD装置541～545の障害発生の有無及び復旧処理中か否かの動作状態を管理する状態情報をHDD装置541～545の状態の変化に応じて書き換え、この状態情報を不揮発性の状態情報記憶部19に格納する。このため、停電等によるシステムダウンが生じても、HDD装置541～545を再立ち上げをするときに、直前状態再現手段7は、この状態情報記憶部19に格納された状態情報に基づいて当該再立ち上げ直前の各HDD装置541～545の動作状態を再現することができる。すなわち、シ

ステムダウンなど不慮の事態が生じても、「正常」、

11

「障害」、「復旧処理中」などの直前の各HDD装置の動作状態が再現される。

【0055】このため、本実施形態では、直前状態再現手段7によって再現された直前状態と、接続状態判定手段8によって判定された現在のHDD装置の接続状態とに基づいて、復旧処理の可否を制御することができる。復旧制御手段9は、直前状態情報及び接続状態情報に基づいて復旧の可否を判定し、例えば、復旧継続制御機能により、直前に行われていた復旧処理を継続させ、または、復旧中止制御機能により直前に行われていた復旧処理を中止させる。

【0056】すなわち、ある実施例では、復旧制御手段9は、直前状態再現手段7によって直前に復旧中であったと判定されたHDD装置54が接続状態判定手段8によって当該直前から現在まで入れ替えられていないと判定されたときには当該復旧を継続させる制御をする復旧継続制御機能(S503→S518)と、直前状態再現手段7によって直前に復旧中であったと判定されたHDD装置54が接続状態判定手段8によって当該直前から現在まで入れ替えられたと判定されたときには当該復旧を中止させる制御をする復旧中止制御機能(S503→S517)とを備える。

【0057】復旧継続制御機能により、直前に復旧処理中であり、また復旧データを記録しているHDD装置と他のHDD装置とが誤って交換されていない場合にのみ自動的に復旧処理を継続することで、他のデータが格納されているHDD装置を誤って接続した場合であっても、この他のデータが格納されているHDD装置のデータに上書きしてしまうことがない。

【0058】また、実施例によれば、復旧制御手段9は、直前状態再現手段7によって直前に復旧中ではないと判定されたHDD装置54が接続状態判定手段8によって当該直前から現在まで入れ替えられたと判定されたときには当該HDD装置54を復旧させるか否かを上位装置へ問い合わせる復旧可否問い合わせ機能(S504→S505)を備える。これにより、新しくHDD装置を障害HDD装置に入れ替えて、復旧可能な場合には、当該HDD装置のデータの復旧の可否を必ず上位装置に問い合わせることとなり、従って、他のデータが格納されたHDD装置が誤って接続されたときであっても、上位装置に確認を求めて、復旧処理によりデータを喪失してしまう事態を有効に防止することができる。

【0059】さらに、実施例によれば、復旧制御手段9は、直前状態再現手段7によって直前に復旧中ではないと判定されたHDD装置54が接続状態判定手段8によって当該直前から現在まで入れ替えられていない判定されたときには当該HDD装置54の動作テストを行うと共に当該動作テストによってエラーが生じなければ当該HDD装置54を復旧させる制御をする復旧開始制御機能(S504→S506)を備える。すなわち、

12

一旦障害が発生したと判定されたHDD装置が交換されずに動作テストでエラーが生じなかった場合には、当該HDD装置に他のデータが格納されていることはないため、自動的に復旧処理を開始する。

【0060】このように、本実施形態では、HDD装置の接続ミスを検出するため、人為的ミスによりデータが失われることがなくなる。さらに、障害HDD装置が交換された時には、人手介入により復旧を開始するため、交換時のミスによりデータが失われることを防げ、復旧を開始した後は、システムがダウンしても再立ち上げ時に自動検出して復旧を開始するため、上位装置を介して入力する必要がなくなる。

【0061】

【実施例】以下、本発明の実施例について図面を参照して詳細に説明する。本実施例では、上述した各手段によりディスクアレイシステムの信頼性を向上させるほか、各記憶部の状態のチェックを行う。具体的には、各記憶部に各種の情報を書き込むときにベリファイを行う。

【0062】すなわち、本実施例では、ディスクアレイコントローラ3が、固有情報を各不揮発性メモリ55に書き込んだときに当該不揮発性メモリ55に書き込んだ固有情報を読み出すと共に当該読み出した固有情報が書き込もうとした固有情報と一致するか否かを確認する固有情報記憶部確認手段を備えている。しかも、状態情報制御手段6が、固有情報記憶部確認手段によって一致しないと判定されたときに当該固有情報のHDD装置54を障害有りと判定する固有情報記憶部障害判定機能を備える。

【0063】また、実施例によれば、固有情報記憶部が、単一の不揮発性メモリを各HDD装置54に割り当てた記憶領域を備える。この場合、固有情報記憶部障害判定機能は、固有情報記憶部確認手段によって記憶領域に格納した固有情報が一致しないと判定されたときには各HDD装置54の全てが障害有りと判定する機能を備える。

【0064】さらに、本実施例では、ディスクアレイコントローラが、状態情報を状態情報記憶部(不揮発性メモリ)19に書き込んだときに当該状態情報記憶部に書き込んだ状態情報を読み出すと共に当該読み出した状態情報が書き込もうとした状態情報と一致するか否かを確認する状態情報記憶部確認手段を備える。

【0065】しかも、状態情報制御手段6が、状態情報記憶部確認手段によって一致しないと判定されたときには各HDD装置54の全てが障害有りであると判定する状態情報記憶部障害判定機能を備える。

【0066】このように、本実施例では、各HDD装置の状態を立ち上げ時に自動検出するとともに不揮発性メモリの障害も検出するため、システムダウン後の再立ち上げ時に誤動作を生じることがなくなる。

【0067】次に、本実施例の動作を説明する。上位装

置61からのアクセス要求（読み出しまだ書き込み要求）は、インターフェース52を介してディスクアレイコントローラ3に入力される。ディスクアレイコントローラ53は、その要求内容に応じて、HDD装置54を制御して、データの読み出しまだ書き込みを行う。また、不揮発性メモリ19内の情報を基にHDD装置の状態の判定を行い、不揮発性メモリ55内の情報とHDD装置の固有情報22を基にHDD装置の状態を判定する。

【0068】本実施例のディスクアレイコントローラ3は、制御用プロセッサとその制御用プロセッサを規定するプログラムが格納されたプログラムメモリと、動作時に使用するパラメータを格納するメモリと、各HDD装置54に読み出しまだ書き込みを行うためのインターフェースと、各HDD装置54にそれぞれ対応した不揮発性メモリ55へのインターフェースと管理情報を格納する不揮発性メモリ19へのインターフェースとにより実現できる。

【0069】〔初期化〕まず、初期化時の動作について説明する。

【0070】本実施例では、HDD装置から読み出して不揮発性メモリ55へ書き込む固有情報D(i)として、ベンダー名とシリアル番号を含む36バイトのInquiryデータを用いている。固有情報はこれに限られるものではなく、たとえば、乱数を用いることや、スロット番号を用いることや、日付や時間を示す情報を用いることができるが、HDD装置ごとに、対応する固有情報が異なるものを使用する。

【0071】図2は固有情報記憶部である不揮発性メモリ1655と、状態情報記憶部である不揮発性メモリ19とが正常か否かを検査する動作例を示すフローチャートであり、図3はこの各メモリ55、19が正常か又は異常と判断されたHDD装置が1台であるときに引き続き初期化処理を行う動作例を示すフローチャートである。この動作は、上位装置からの初期化指示により開始される。

【0072】初期化指示を受信したディスクアレイコントローラ53は、各HDD装置を順次特定する変数iに”1”、障害HDD装置台数カウンタNEERに”0”、障害と判定したHDD装置を特定する管理情報配列C(i)に”0”をセットする(S101)。

【0073】カウンタiはHDD装置を識別するためのカウンタであり、障害HDD装置台数カウンタN\_EERは、使用不可能なHDD装置の台数のカウントに用いられ、管理情報配列C(i)では、使用不可とするHDD装置に”1”をセットする。

【0074】これらのパラメータの格納には、ディスクアレイコントローラ3内のメモリが用いられる。この後、ディスクアレイコントローラ3は管理情報格納用の不揮発性メモリ19が正常であるか否かのテストを、ベ

リファイにより行う(S102)。

【0075】管理情報格納用の不揮発性メモリ19が正常でない場合(N)には、ディスクアレイシステム障害としてエラーメッセージを上位装置に出力し(S120)、初期化動作を終了する。

【0076】管理情報格納用の不揮発性メモリ19が正常である場合(S102:Y)には、i番の不揮発性メモリ55が正常であるか否かのテストをベリファイにより行う(S103)。

【0077】i番の不揮発性メモリ55が正常である場合(S103:Y)には、S106へ進み、i番の不揮発性メモリ55が正常でない場合(S103:N)には、NEERに”1”を加算、C(i)に”1”をセットし、管理情報を格納するための不揮発性メモリ19へC(i)を書き込む(S104、初期化制御手段5)。すなわち、固有情報記憶部である不揮発性メモリ55が異常である場合には、当該不揮発性メモリ55に併設されたHDD装置を障害有りとする。

【0078】この不揮発性メモリ19への管理情報C(i)の書き込みが正常でない場合(S105:N)には、ディスクアレイシステム障害としてエラーメッセージを上位装置に出力(S120)して処理を終了する。すなわち、状態情報記憶部である不揮発性メモリ19が異常である場合には、ディスクアレイシステム全体を障害有りとする。一方、不揮発性メモリ19への書き込みが正常である場合(S105:Y)には、S106へ進む。

【0079】S106では、パラメータiに”1”が加算され、iが全HDD装置の台数以下の場合(S107:Y)には、S103へ戻り、次の不揮発性メモリ55のテストを行う。変数iが全HDD装置台数より大きくなつたとき(S107:N)には、全ての不揮発性メモリ55の検査が終了したため、NEER値の判定(S108)を行う。

【0080】NEERが”2”以上である場合(S108:Y)には、1台のHDD装置のデータを復旧させるシステムでは、ディスクアレイシステムの動作が不可能となるので、ディスクアレイシステム障害としてエラーメッセージを上位装置に出力し(S120)、初期化動作を終了する。NEERが”1”以下である場合(S108:N)には、磁気ディスク装置としての動作が可能であるため、図3のS109へ進む。

【0081】図3では、S109ではカウンタiに”1”、固有情報配列D(i)に”0”をセットする。固有情報配列D(i)は、各HDD装置54に予め格納された固有情報を格納する配列である。この固有情報配列D(i)は、再立ち上げ処理において、各固有情報記憶部である不揮発性メモリ55に格納された固有情報の配列E(i)と比較される。

【0082】i番の不揮発性メモリ55が正常でなかつ

た場合 (S110:Y) は、S117へ進む。i番の不揮発性メモリ55が正常だった場合 (S110:N) には、i番のHDD装置からHDD装置の固有情報D

(i) を読み出し (S111) 、読み出しが正常に行えなかった場合 (S112:N) には、NEERに”1”を加算、C(i)に”1”をセットし、管理情報を格納するための不揮発性メモリ19へC(i)を書き込み (S115) 、書き込みが正常でない場合 (S116:N) には、ディスクアレイシステム障害としてエラーメッセージを上位装置に出力 (S120) して処理を終了し、書き込みが正常である場合 (S116:Y) には、S117へ進む。

【0083】i番のHDD装置のD(i)が正常に読み出せたとき (S112:Y) には、i番の不揮発性メモリ55にD(i)を書き込み (S113) 、書き込みが正常に行えなかった場合 (S114:N) には、NEERに”1”を加算、C(i)に”1”をセットし、管理情報を格納するための不揮発性メモリ19へC(i)を書き込む (S115) 。

【0084】この不揮発性メモリへの書き込みが正常でない場合 (S116:N) には、ディスクアレイシステム障害としてエラーメッセージを上位装置に出力 (S120) して処理を終了し、書き込みが正常である場合 (S116:Y) には、S117へ進む。

【0085】i番の不揮発性メモリ55へD(i)の書き込みが正常に行えた場合 (S114:Y) には、S117へ進む。S117では、パラメータiに”1”が加算され、iが全HDD装置の台数以下の場合 (S118:Y) には、S110へ戻り、次のHDD装置からの固有情報の読み込みと不揮発性メモリ55への書き込みが行われる。

【0086】iが全HDD装置台数より大きくなったり (S118:N) に、NEER値の判定 (S119) が行われる。NEERが”0”である場合には、全HDD装置および不揮発性メモリ55が正常であるので、初期動作を完了させる。

【0087】NEERが”2”以上の値であるときには、磁気ディスク装置の動作が不可能となるので、障害磁気ディスク装置としてのエラーメッセージを上位装置に出力 (S120) して、初期化動作を終了する。障害磁気ディスク装置と判断された磁気ディスク装置は、記憶装置として動作させることはできない。

【0088】また、NEERが”1”であるときには、データをデータを冗長構成とすることはできないが、記憶装置として動作させることは可能であるので、C(i)が”1”的HDD装置を使用不可として (S121) 、初期化動作を終了する。

【0089】この図3に示す初期化処理は、初期化制御手段5により制御される。

【0090】[障害発生時] 次に、ディスクアレイシス

テムが記憶装置として動作しているときに、HDD装置541～545のうちの1台に障害が発生した場合の動作を説明する。ディスクアレイコントローラ3は、個々のHDD装置で障害が発生した場合、不揮発性メモリ19の管理情報の変更を行う。具体的には、このディスクアレイシステムでは、障害が発生したHDD装置に対応する管理情報のビットを”1”に書き換える。

【0091】管理情報として他の形式の方法を用いた場合には、この変更方法もそれに応じて修正される。たとえば、正常時には管理情報を示すビットが全て”1”とする場合には、障害が発生すると、管理情報のビットを”0”にするように装置を構成すればよい。

【0092】[復旧処理時] また、本実施例では、不揮発性メモリ19へ書き込むHDD装置の状態を示す管理情報C(i)として、障害HDD装置を示すビットと復旧中のHDD装置であることを示すビットをそれぞれのHDD装置に対応させて用いている。HDD装置が全て正常に動作しているときは、障害HDD装置を示すビットは全て”0”であり、障害HDD装置がある場合はその障害HDD装置にに対応するビットは”1”である。

【0093】復旧を行っているHDD装置は、復旧中のHDD装置を示すビットが”1”であるが、このとき復旧中を示すビットが”1”になっているものは1つだけである。本実施例では、この復旧中であるか否かを示す情報に基づいて、再立ち上げ処理を制御する。

【0094】[再立ち上げ] 本実施例のディスクアレイシステムでは、不揮発性メモリ19の管理情報と、HDD装置の固有情報と不揮発性メモリ55に格納されている固有情報を基に、装置の再立ち上げが行われたときは、以下のような手順で、各HDD装置の状態の判定を行う。

【0095】図4乃至図8に、再立ち上げ時のディスクアレイコントローラ53の動作の流れを示す。図4乃至図8に示す処理では、符号A, B, C, D, E, Fでそれぞれ連続している。

【0096】まず、カウンタiに”1”、障害HDD装置台数NEERに”0”、障害HDD装置識別パラメータNDISKに”0”、不揮発性メモリ19に格納されている障害状態を示す管理情報配列C(i)に”0”、各HDD装置から読み出した固有情報の配列D(i)に”0”、各不揮発性メモリ55から読み出した固有情報の配列E(i)に”0”、不揮発性メモリ19に格納されている復旧状態を示す管理情報配列F(i)に”0”をセット (S401) する。

【0097】そして、不揮発性メモリ19から、各HDD装置の状態を示す管理情報配列C(i)と、復旧状態を示す管理情報配列F(i)とを読み出す (S402) 。読み出しが正常に行えなかった場合 (S403:N) には、ディスクアレイシステムの障害としてエラーメッセージを出力し (S516) 、処理を終了する。

【0098】読み出しが正常に行えた場合 (S403 : Y) には、C (i) が”1”であれば (S404 : Y) 、図5に示すS415に処理を移行し、i番のHDD装置から固有情報を読み出して、D (i) に格納する (S415)。読み出しが正常でなければ (S417 : N) S420へ進む。

【0099】一方、読み出しが正常であれば (S417 : Y) 、i番の不揮発性メモリからE (i) を読み出し (S416)、この読み出しが正常でなければ (S418 : N) 、S420へ進み、読み出しが正常であれば (S418 : Y) S412へ進む。

【0100】S420では、i番のHDD装置から読み出した固有情報D (i) とi番の不揮発性メモリから読み出した固有情報が異なるように、固有情報の一部の書き換えをおこない、S412へ進む。このステップS420での固有情報の書き換えにより、i番の不揮発性メモリ55に異常があることを記憶する。S412では、NEERに1を加算、N<sub>DISK</sub>にiをセットし、図6に示すS413へ進む。

【0101】図4に示すステップS404において、C (i) が1でなければ (S404 : N) 、すなわち、再立ち上げ直前に障害なしであったHDD装置については、図6に示すステップS405に処理を移行し、このi番のHDD装置から固有情報D (i) を読み出す (S405)。読み出しが正常に行えなかった場合 (S406 : N) には、S419へ進み、読み出しが正常に行えた場合 (S406 : Y) には、i番の不揮発性メモリ55から格納されている固有情報E (i) を読み出す (S407)。固有情報E (i) の読み出しが正常に行えなかった場合 (S408 : N) には、S419へ進む。

【0102】S419では、HDD装置の異常又はその不揮発性メモリ55に異常がある場合には、i番のHDD装置から読み出した固有情報D (i) とi番の不揮発性メモリから読み出した固有情報が異なるように、固有情報の一部の書き換えをおこない、S410へ進む。

【0103】固有情報E (i) の読み出しが正常に行えた場合 (S408 : Y) には、i番のHDD装置から読み出した固有情報D (i) とi番の不揮発性メモリ55から読み出した固有情報E (i) を比較し、D (i) とE (i) が異なる場合 (S409 : N) にはS410へ進み、D (i) とE (i) が等しい場合 (S409 : Y) には、S413へ進む。このS409で一致しない場合には、そのHDD装置が交換されたことを意味する。

【0104】このため、S410ではNEERに”1”を加算、N<sub>DISK</sub>にiをセット、C (i) に”1”をセットし、管理情報を格納するための不揮発性メモリ19へC (i) を書き込み (S410)、書き込みが正常でない場合 (S411 : N) には、処理2を行い、書き込みが正常である場合 (S411 : Y) には、S413へ進

む。

【0105】さらに、S413では、パラメータiに”1”が加算され、iが全HDD装置の台数以下の場合 (S414 : Y) には、図4のS404へ戻り、次のHDD装置の固有情報と不揮発性メモリ55の固有情報との比較が行われる。

【0106】このような動作をiが全HDD装置台数より大きくなるまで、繰り返すことにより (S414 : N) 、NEERには正常でないHDD装置の数がセットされ、管理情報C (i) には、障害の発生しているHDD装置を識別するビットがセットされることになる。

【0107】また、管理情報を格納する不揮発性メモリ55の障害によるディスクアレイシステム障害が検出されたことになる。S419およびS420では、D (i) に”FFh”、E (i) に”00h”を書き込んでいるが異なるようにすれば良いので値はこれに限らない。これは、読み出しには失敗するがデータは正常なものが送られてくる場合があるので、回路の設計方法やデータの格納方法によっては使用しなくても構わない場合もある。

【0108】この図4乃至図6に示す処理により、現在のHDD装置の接続状態が確認され、交換されている場合には、その交換されたHDD装置の番号がN<sub>DISK</sub>に格納された。この接続状態の確認は、接続状態判定手段8により行われる。

【0109】次いで、図7及び図8に示す処理を行う。まず、障害HDD装置台数カウンタNEERの判定を行い (S501)、NEERが”0”であるときには、全HDD装置が正常であるので、立ち上げ動作を終了する。

【0110】NEERが”2”以上であるときには、実施例のディスクアレイシステムの冗長構成ではカバーできない障害であるため、障害磁気ディスク装置とし、エラーメッセージを出力して (S516)、処理を終了する。

【0111】NEERが”1”であるときには、復旧状態を示す管理情報F (N<sub>DISK</sub>) から電源切断前の状態が復旧中であるかを判断する (S502)。電源切断前の状態が復旧中であれば (S502 : Y) 、HDD装置から読み出した固有情報D (N<sub>DISK</sub>) と不揮発性メモリ55から読み出した固有情報E (N<sub>DISK</sub>) が同じであるか比較する (S504)。この比較により、F (N<sub>DISK</sub>) で特定されるHDD装置が交換されていないと判定された場合には、ステップS518以下の自動復旧処理を行う。一方、異なれば (S503 : N) 当該HDD装置を使用不可とする (S517)。

【0112】S502において、電源切断前が復旧中でない場合には、HDD装置から読み出した固有情報D (N<sub>DISK</sub>) と不揮発性メモリ55から読み出した固有情報E (N<sub>DISK</sub>) が同じであるか比較し、同じであれば (S504 : Y) 、一旦障害有りと判定されたHDD装

置であるため、再度このN<sub>DISK</sub>番のHDD装置が正常か否かのテストを行う(S506)。そして、正常であれば(S507:Y)、復旧状態を示す管理情報配列F(N<sub>DISK</sub>)に1を格納し、自動復旧処理を行う(S518)。一方、N<sub>DISK</sub>番のHDD装置のテストが正常でなければ(S507:N)、当該HDD装置を使用不可能とする(S517)。

【0113】また、S504において、HDD装置から読み出した固有情報D(N<sub>DISK</sub>)と不揮発性メモリ55から読み出した固有情報E(N<sub>DISK</sub>)が同じであるか比較し、異なる場合(S504:N)には、すなわち、障害処理中ではないHDD装置が交換された場合には、図8に示すステップS505に処理を移行し、上位装置からの復旧命令を待つ(S505)。そして、復旧命令を受信したら(S505:Y)、N<sub>DISK</sub>番のHDD装置のテストを行い(S508)、HDD装置のテストの結果が正常でない場合(S509:N)には、使用不可とし(ステップS517)、HDD装置のテストが正常な場合(S509:Y)には、N<sub>DISK</sub>番のHDD装置からD(N<sub>DISK</sub>)を読み出す(S510)。

【0114】読み出しが正常に行えなかった場合(S511:N)にはS517へ進み、読み出しが正常に行えた場合(S511:Y)には、N<sub>DISK</sub>番の不揮発性メモリ55へD(N<sub>DISK</sub>)を書き込む(S512)。

【0115】不揮発性メモリ55へD(N<sub>DISK</sub>)が正常に書き込めなかった場合(S513:N)には、S517へ進み、正常に書き込めた場合(S513:Y)には、復旧状態を示す管理情報F(N<sub>DISK</sub>)に”1”をセットして、不揮発性メモリ19へF(N<sub>DISK</sub>)を書き込む(S514)。

【0116】書き込みが正常に行えなかった場合(S515:N)には、処理2へ進み、書き込みが正常に行えた場合(S515:Y)には、復旧作業を開始する(S518)。

【0117】図7及び図8に示すS518では、はN<sub>DISK</sub>番のHDD装置を他のHDD装置内のデータから再構築されたデータを書き込む復旧作業を行い(S518)、復旧作業が終了したら、管理情報C(N<sub>DISK</sub>)、F(N<sub>DISK</sub>)にそれぞれ”0”をセット(S519)して再立ち上げ動作を終了する。

【0118】以上説明した実施例の磁気ディスク装置では、固有情報を格納する不揮発性メモリを個々のHDD装置に対応して備え、それとは別に管理情報を格納する不揮発性メモリを用いた構成をしているが、固有情報を格納する不揮発性メモリを1つにして、個々のHDD装置の固有情報を格納する領域に分けた場合や、管理情報と固有情報を全て1つの不揮発性メモリに領域を分けて格納する場合があるが、このときは、S103のN、S113のNの場合がS120へ、S408のN、S418のNが処理2へ、S512のNがS516へ行くよう

に変更する必要がある。

【0119】これは固有情報を格納する不揮発性メモリを1つにした場合、1つの領域の読み出し、または書き込みに失敗するということは、その不揮発性メモリ自体の動作がおかしいことになるためであり、1つの不揮発性メモリの他の領域の読み込み、または書き込みが成功してもそのデータの信頼性が著しく低下するため、この不揮発性メモリを使用しないようにするためである。

【0120】実施例の磁気ディスク装置では、1台のHDD装置の復旧しか行えない冗長構成を採用しているが、障害HDD装置識別パラメータを復旧可能な台数に応じて増やし、障害HDD装置台数カウンタによる分岐条件をすることで、2台以上のHDD装置を復旧できる冗長構成をとることができる。

【0121】また、実施例の磁気ディスク装置では、立ち上げ時に一部自動的に復旧可能なディスクに対しては復旧を行うように構成してあるが、この機構を設げずに、上位装置からの指示により、復旧作業が開始されるように構成してもよい。

【0122】以上説明したように、本実施例によると、各HDD装置の状態を立ち上げ時に立ち上げ時に自動検出するとともに不揮発性メモリの障害も検出するため、システムダウン後の再立ち上げ時に誤動作を生じることがない。また、HDD装置の接続ミスも検出するため、人為的ミスによりデータが失われることがない。さらに、障害HDD装置が交換された時に、人手介入により復旧を開始するため、交換時のミスによりデータが失われることを防げ、復旧を開始した後は、システムがダウンしても再立ち上げ時に自動検出して復旧を開始するため、上位装置を介して入力する必要がない。

### 【0123】

【発明の効果】本発明は以上のように構成され機能するので、これによると、状態情報制御手段が、磁気ディスク装置(HDD装置)の障害発生の有無及び復旧処理中か否かの動作状態を管理する状態情報(C(i), F(i))を磁気ディスク装置の状態の変化に応じて書き換え、復旧制御手段が、接続状態判定手段によって判定された接続状態情報(D(i), E(i))と直前状態再現手段によって再現された直前の各磁気ディスク装置の動作状態情報(C(i))とに基づいて復旧処理の継続又は開始を制御するため、立ち上げ前の直前の状態と、現在の接続状態とに応じて、復旧処理を行うか否かを定めることができ、例えば、人為的な接続ミスがない場合にのみ自動的に復旧処理を開始し、また、交換がされたときには上位装置へ問い合わせを行うことができるため、復旧処理により必要なデータを上書きすることなく、従って、信頼性をより向上させることができる従来にない優れたディスクアレイシステムを提供することができる。

【図面の簡単な説明】

【図1】本発明の一実施形態の構成を示すブロック図である。

【図2】本実施例による初期化処理の前段を示すフローチャートである。

【図3】本実施例による初期化処理の図2に続く後段を示すフローチャートである。

【図4】本実施例による立ち上げ処理での立ち上げ直前のHDD装置の障害状態によって処理を分岐する動作の一例を示すフローチャートである。

【図5】本実施例による立ち上げ処理での立ち上げ直前に障害HDD装置であった場合の図4に続く接続状態確認処理の一例を示すフローチャートである。

【図6】本実施例による立ち上げ処理での立ち上げ直前に諸障害HDD装置ではなかった場合の図4に続く接続状態確認処理の一例を示すフローチャートである。

【図7】図5及び図6に示す接続状態確認処理に続く再立ち上げ動作の一例を示すフローチャートである。

【図8】図7に示す立ち上げ処理で直前に復旧処理が行われておらずかつHDD装置が交換された場合の処理を示すフローチャートである。

【図9】従来例のRAIDのレベル0の磁気ディスク装置の書き込み動作の概要を示す説明図である。

【図10】従来例のRAIDのレベル4のディスクアレイシステムの書き込み動作の概要を示す説明図である。

【図11】従来のディスクアレイシステムの概略構成の一例を示すブロック図である。

【図12】従来例のディスクアレイシステムの立ち上げ時における各HDD装置の状態判定を行う動作のフロー

チャートである。

【図13】従来例の磁気ディスク装置の立ち上げ時に、図12の流れ図に引き続いて行われる動作を示すフローチャートである。

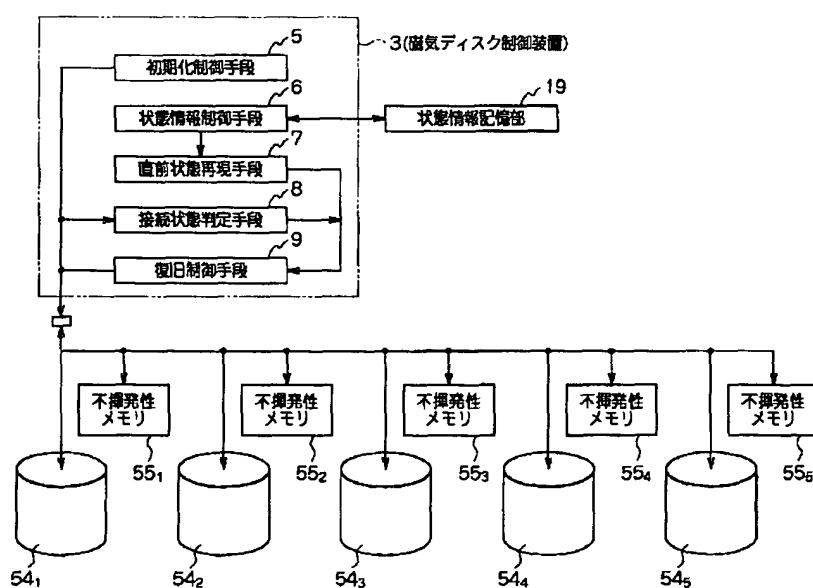
【図14】従来例の構成を2つ以上持つディスクアレイシステムの障害HDD装置交換例を示す説明図である。

#### 【符号の説明】

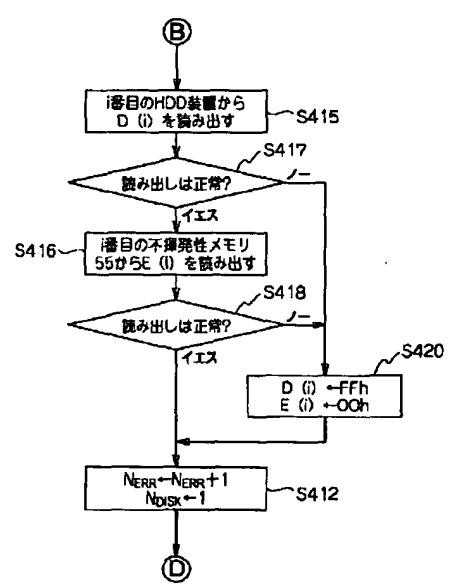
3, 53, 53a, 53b 磁気ディスク制御装置（ディスクアレイコントローラ）

- 10 5 初期化制御手段
- 6 状態情報制御手段
- 7 直前状態再現手段
- 8 接続状態判定手段
- 9 復旧制御手段
- 19 状態情報記憶部（不揮発性メモリ）
- 51, 51a, 51b ディスクアレイシステム
- 52, 52a, 52b インターフェース
- 54, 541～545, 54a1～54a5, 54b1～54b5 磁気ディスク装置（HDD装置）
- 20 55, 551～555, 55a1～55a5, 55b1～55b5 固有情報記憶部（不揮発性メモリ）
- 56, 56a, 56b 時計回路
- 57, 571～575, 57a1～57a5, 57b1～57b5 管理情報記録領域
- 581～585 記録領域
- 60 書き込み要求データ
- 61 上位装置

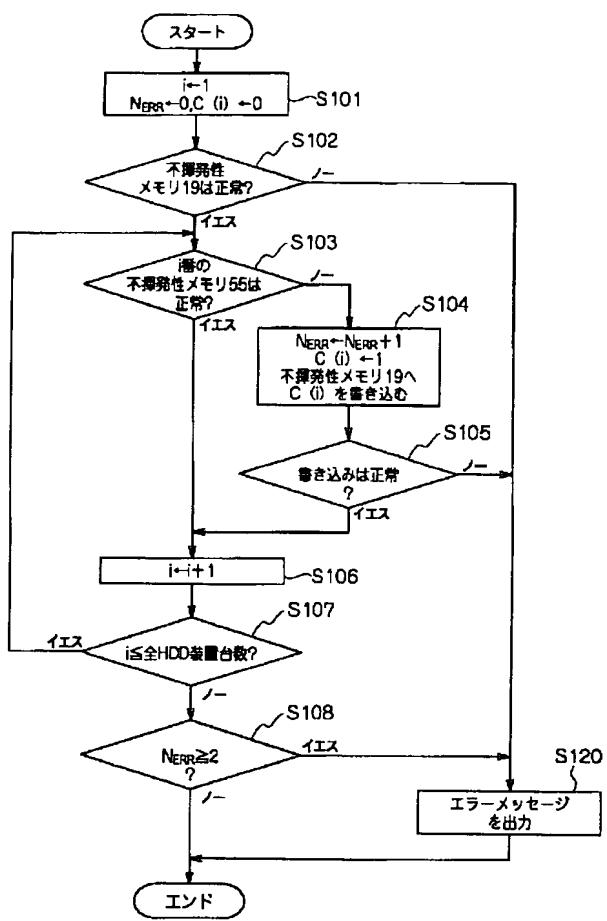
【図1】



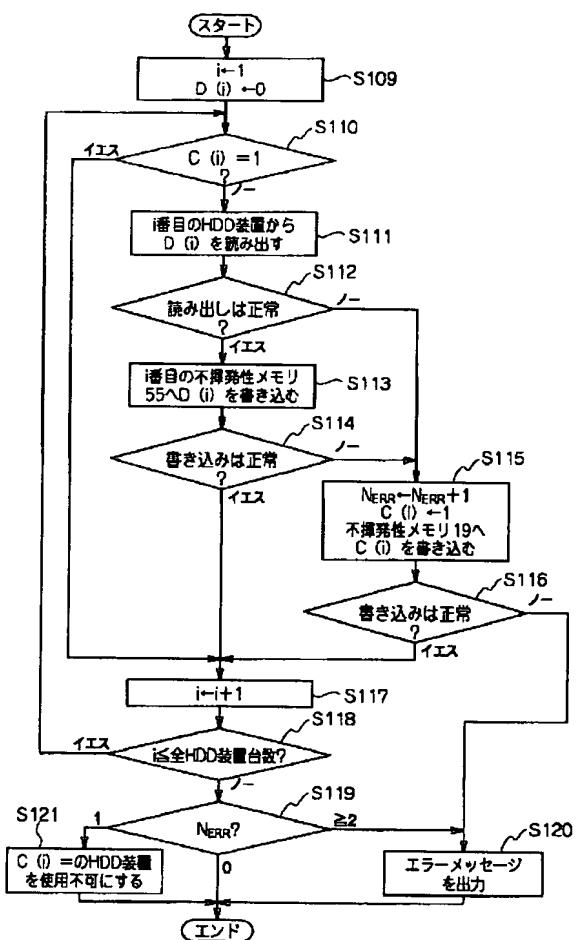
【図5】



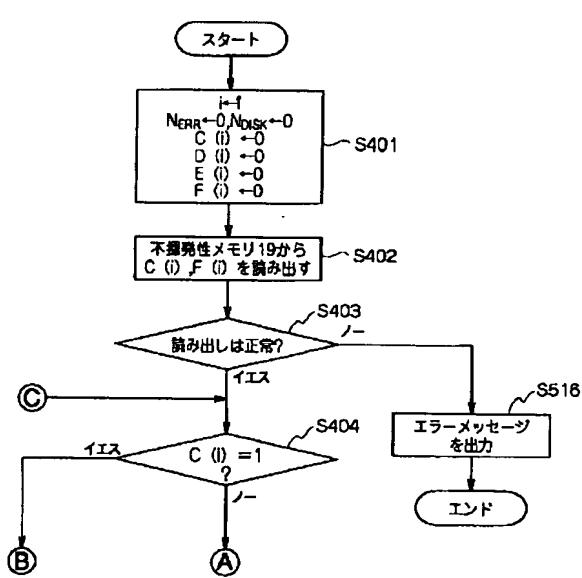
【図2】



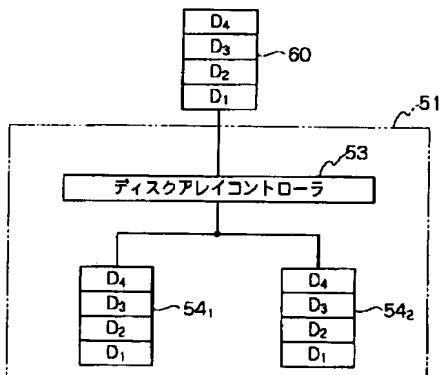
【図3】



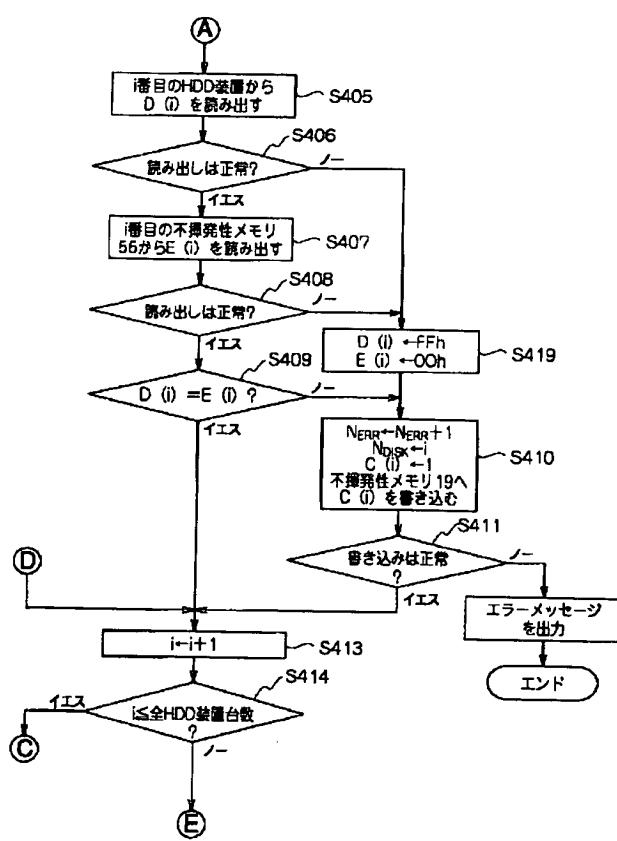
【図4】



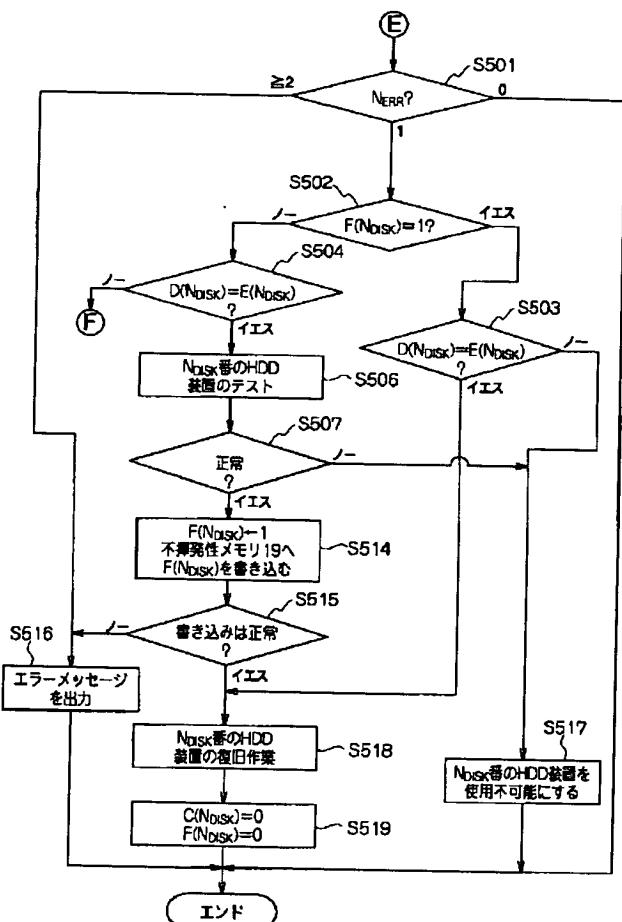
【図9】



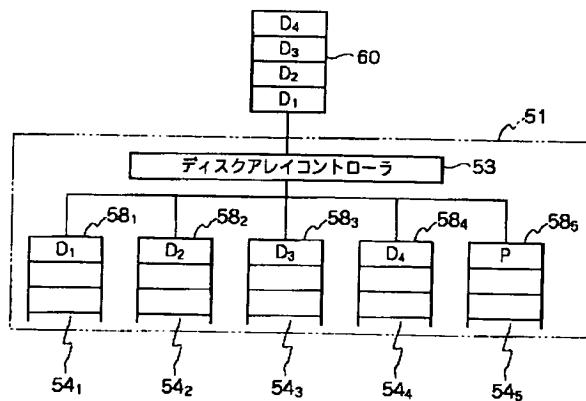
【図6】



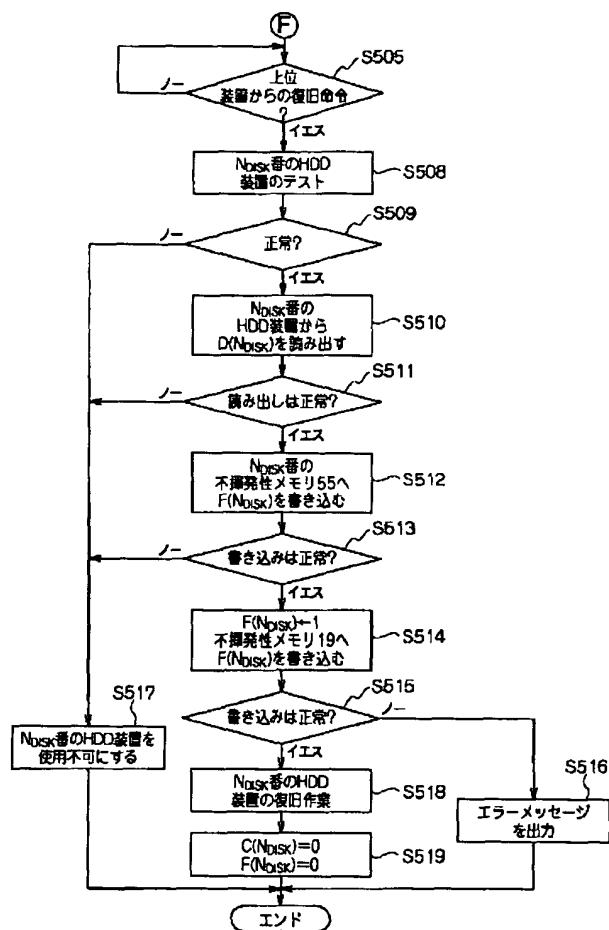
【図7】



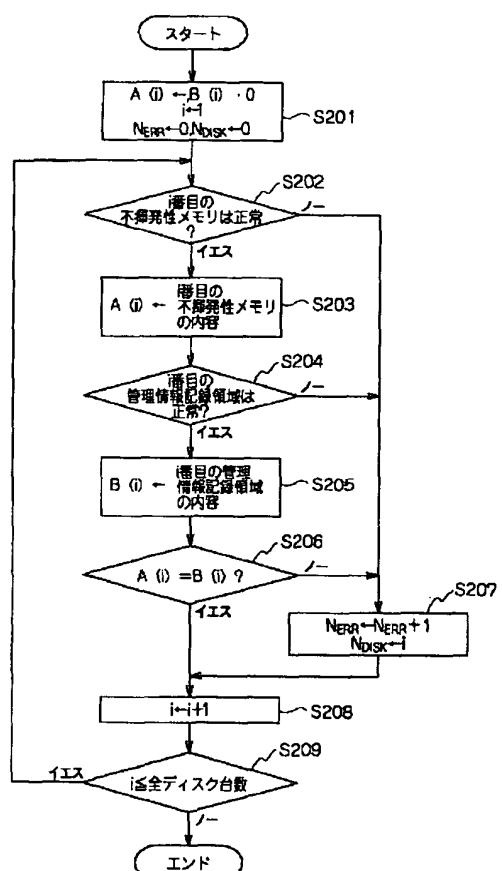
【図10】



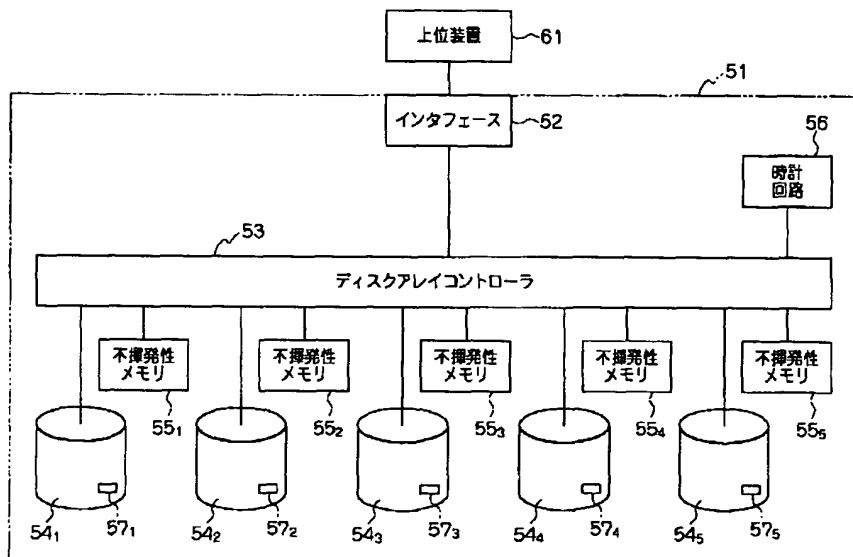
【図8】



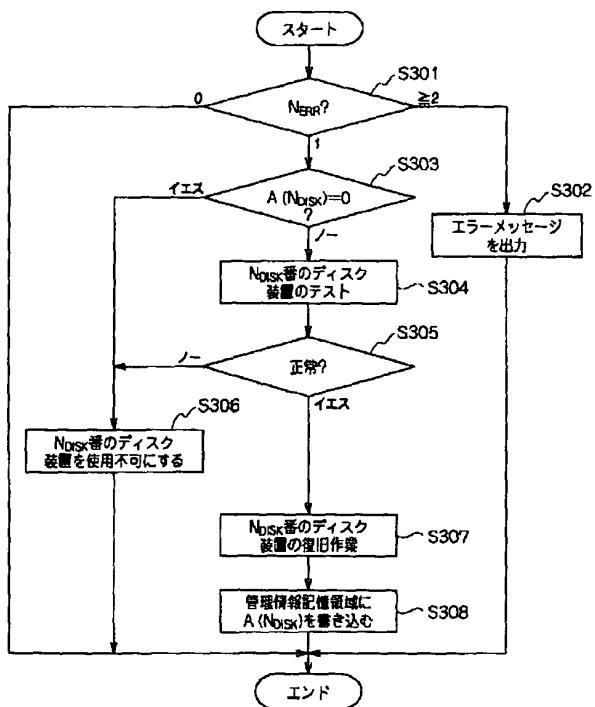
【図12】



【図11】



【図13】



【図14】

